Clustering Wireless Ad Hoc Networks with Boundary Nodes

Chih-Cheng Tseng*
Graduate Institute of Communication Engineering,
National Taiwan University,
Taipei, Taiwan, R.O.C.
cctseng@mail.jwit.edu.tw

Kwang-Cheng Chen
Graduate Institute of Communication Engineering,
National Taiwan University,
Taipei, Taiwan, R.O.C.
chenkc@cc.ee.ntu.edu.tw

Abstract—**Based on the observation that most of the orphan clusters are generated from boundary nodes, we propose a Boundary-First Cluster-Minimized (BFCM) clustering algorithm to minimize the number of generated orphan cluster by boundary node. The proposed algorithm is compared to the well known ID-based and Degree-based algorithms. The simulation results show that with the same message and time complexities as the compared algorithms, the proposed algorithm generates the minimum number of orphan clusters and, thus, the total number of generated clusters are minimized.**

Keywords-ad hoc networks, boundary node, cluster, orphan cluster, typical node

I. INTRODUCTION

Wireless ad hoc network is a self-organizing network architecture that can be rapidly deployed and can also be dynamically adapted to the propagation and the traffic conditions and the mobility patterns of the wireless nodes. Possible examples of the wireless ad hoc networks are tactical military applications, disaster recovery operation and exhibitions or conferences. The most distinguishing characteristic of the wireless ad hoc networks is the lack of the fixed infrastructure. Thus, designing an efficient and stable operational architecture turns out to be an important issue. One of the general approaches is to partition the entire network into groups of *clusters*. Within each cluster, a node is elected as a *clusterhead* to control communications among the cluster. Some advantages to organize the entire network into clusters are listed as follows.

Frequency spatial reuse: With clustering, the channel assignment strategy can be employed to optimally and spatially reuse the radio frequency among clusters.

Power consumption: In order to achieve mobility, power of the mobile device is mainly supplied by batteries. To prolong the communication duration of the device, the transmission power must be efficiently utilized in order to conserve the limited battery power.

Interference: Without clustering, higher transmission power is needed in order to achieve the global network connectivity. As a result, interference to and from neighboring nodes will be very severe. Through partitioning the entire network into clusters, the interference is reduced since only local connectivity is required.

Robustness: In a fully distributed wireless ad hoc network, it is very difficult for each wireless node to have the correct knowledge about the entire network topology due to the random mobility of the wireless nodes. By clustering, however, each wireless node only needs to know the local network topology information.

Increase system capacity: With clustering, the capacity can be improved both because the frequency reuse and the reduction in transmitting network topology maintenance messages and routing information.

In this paper, to take the effects of boundary nodes into account, we propose a distributed Boundary-First Cluster-Minimized (BFCM) clustering algorithm to reduce the total number of generated clusters by minimizing the number of organized clusters. We show that with this algorithm, the numbers of the organized clusters are reduced dramatically with compared to the well-known ID-based and Degree-based algorithms. The rest of the paper is organized as follows: in Section II, a summary of related work is presented. Section III formulates the cluster formation problem as a cluster minimization problem and proposed the BFCM clustering algorithm to achieve the objective. The simulation results and discussions are presented in Section IV, and Section V concludes this paper.

II. RELATED WORK

Many related algorithms are proposed in the literature. The *Minimum connected dominating set (MCDS)* approach [1]-[2] is inspired by the graph theory. In this approach, a connected graph $G = (V, E)$ is used to represent a wireless ad hoc network, where V represents a set of wireless nodes and E represents a set of edges. A subset of the vertices in a graph is a *dominating set* if every vertex not in the subset is adjacent to at least one vertex in the subset. A *connected dominating set* C of a graph G is a connected subgraph of G such that any vertex in C can communicate with each other without using vertices in V-C. In facts, nodes within the minimum connected dominating set and edges connect nodes within the set form a virtual backbone. Then, the clusterheads and the gateways can be easily elected from the virtual backbone. The main advantage of the MCDS scheme is to obtain an optimum configuration as the backbone of the wireless ad hoc networks in such a way, as long as the network topology changes do not affect the backbone there is no need to reconstruct the network. How-

[*] C-C Tseng is also with the Department of Electronic Engineering, Jin-Wen Institute of Technology, Xin-Dian, Taiwan, R. O. C.

ever, the difficulty is that the problem to find a minimum connected dominating set in a connected graph is shown to be NP-hard [3] and the problem to find the optimal clusterhead set is NP-complete [4]. The general feasible alternative is to find an approximated heuristic algorithm to achieve a sub minimum connected dominating set. In [5], a Degree-based approach is proposed to select nodes with the highest degree as clusterhead, while the ID-based scheme [6]-[7] organized the cluster simply based on the node ID. One of the well-known ID-based approaches to organize the wireless ad hoc wireless networks is the linked cluster algorithm (LCA) [6] in which nodes are organized into set of clusters with each node belonging to at least one cluster. Within each clusters, a node is elected to be clusterhead and acts as a local controller of the cluster. For inter-cluster communications, gateway nodes are nodes at the fringe of a cluster and can communicate to other clusters. The advantage of the ID-based scheme is insensitive to the network dynamics compared to the Degree-based. The disadvantages of this scheme are: (1) Node needs to periodically broadcast the list of nodes that it can hear (including itself). This wastes large precious bandwidth when the number of node increased. (2) The election method for clusterheads and gateway nodes critically biased to those nodes with specific IDs. Shah and Flikkema [8] select a clusterhead that take link losses and transmitter power into consideration by using a link loss matrix. Singh and Kurose [9] propose a delay model for each node to calculate the total delay from every other node to it and, then, the node with the minimum delay is considered as more "central" and would be a better choice for leader than other nodes. Basagni [10] associate the speed and power of a node as the weighting factor; the slower of the speed and the stronger of the received power, the better of the node being elected as clusterhead. In [11], clusterhead selection is based on the ratio of power level due to successive receptions at each node from its neighbors. Results show that it achieves better performance than the ID-based algorithms.

In viewing these previous works, we find that *stability* to the dynamical change of network topology is the main design consideration in their algorithms. However, as stated in [12], the more number of clusters that a clustering algorithm generated, the more number of inter-cluster and intra-cluster information exchanges needed. As a consequence, the overall available bandwidth will be wasted in exchanging network information. Thus, our objective is to design an efficient clustering algorithm to minimize the number of cluster generated so that the available bandwidth is maximized.

III. THE CLUSTER MINIMIZATION ALGORITHM

A. NETWORK MODEL AND DEFINITIONS

The entire network is modeled as a connected undirected graph $G(V,E)$, where V is the set of nodes with cardinality N, i.e., $V=\{1,2,\ldots,N\}$ and E is the set of edges $E=\{(i,j):i,j\in V\}$. Every node v in the set V is assigned a unique ID, denoted by the numbers 1,2…, N, where N is the number of nodes in the network. Furthermore, it is assumed that edges are all bi-directional and the transmission range for each node is fixed and identical. Thus if (i,j) is an edge between node i and node j, then so is (j,i) between node j and node i. We also assume that signaling packets and data packets are exchanged over a common error-free wireless channel. We define following terminologies: A node in a cluster is selected as *clusterhead* if it is center located. Node that is not clusterhead is called *ordinary node*. All nodes are assumed to be identical, that is, no node with additional capabilities such that it tends to be easily elected as a clusterhead. A node is *marked* if it belongs to some cluster; otherwise, it is *unmarked*. All nodes are assumed unmarked initially. The *degree* of a node is the number of one-hop neighbors that the node connects with. A node with degree 1 is said to be a *boundary node* and the only neighbor of a boundary node is said to be a *typical node*. A cluster is said to be an *orphan cluster* if it contains only one node and degree of this node is greater than 0.

B. PROBLEM FORMATION

As described in Section II, when supporting multimedia services over wireless ad hoc networks, an important design requirement arise: the number of constructed clusters. This will incur some severe issues. First of all, from routing point of view, the more the number of clusters organized, the more routing overheads are needed to maintain the organized topology. This will result in reducing the total available system capacity. Second, with clusterhead in each cluster, when two direct connected nodes proceed to have an intra-cluster or inter-cluster communications, they are required to go through the clusterhead. Thus, we can only construct a sub-optimal route fro the source to the destination. From QoS admission control point of view, the more the number of organized clusters, the longer the distance for a source node to reach its destination and will incurs the larger delay and higher delay jitter. Thus, designing a clustering algorithm that can organize the network topology into an efficient architecture to support QoS guaranteed multimedia services is the main consideration of this paper. Therefore, we formulate this clustering problem as follows.

Objective:

Minimize the total number of organized clusters.

Constraints:

(*dominance* constraint) Every ordinary node has at least a clusterhead as neighbor.

(*independence* constraint) There is no overlap between clusters.

To this minimization problem, we find that the objective to minimize the number of organized clusters can be transformed into two sub-objectives: maximizing the number of nodes inside a cluster and minimizing the number of orphan clusters. For the first sub-objective, we define a *greedy* clusterhead election criterion: the more neighbors of a node connect with, the more likely for the node to be elected as a clusterhead. By this criterion, the number of node within a cluster will be maximized and, therefore, the number of organized clusters

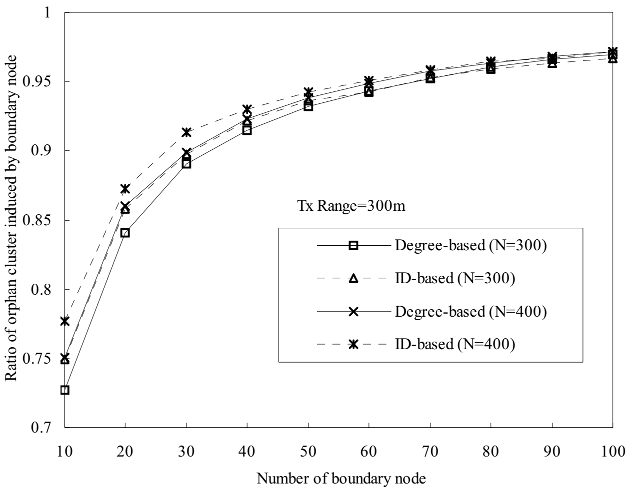

Figure 1 The ratio of orphan clusters that are organized by boundary node.

will be minimized. As to the second sub-objective, since the role of each node plays in the network is determined after negotiated with neighbors, it is difficult to characterize if a node will turn out to be organized into an orphan cluster before clustering. To solve this difficulty, we exam the cluster architectures created by the well-known ID-based and Degree-based algorithms. An important characteristic is found that more than 70% of orphan clusters are generated by boundary nodes as depicted in Figure 1. The main reason for this is that if the only neighbor of a boundary node, i.e. typical node, joins into a cluster constructed by its neighbor (other than the boundary node), with no other choice, this boundary node definitely organizes an orphan cluster. This important characteristic gives us an idea to reduce the generation of orphan cluster: always starting clustering from the boundary node if it exists, which might be realized by two alternatives. The first one is to select boundary node as clusterhead and the second one is to select the corresponding typical node. In the former case, the organized cluster contains only two nodes; however, the number of nodes within a cluster in the later case is more than two (depends on the degree of typical node). Combining with the first sub-objective, we select the later alternative to design our clustering algorithm. To satisfy the above two sub-objectives, we define a generalized weighting function to assign weighting of a node v, $w(v)$, as

$$w(v)=\begin{cases}\Delta, & \text{if } v \text{ is a typical node}\\ f(v), & \text{otherwise}\end{cases} \quad (1)$$

where $f(v)$ is function that maps node attributes that are considered into weighting value and Δ is parameter that guarantees typical node to be the maximum weighting node in the network no matter what node attributes are considered. For example, for a connected ad hoc network with N nodes, if node ID is regarded as the weighting, i.e. BFCM-ID algorithm, then $f(v)=N-node\_id$ and Δ may be set to the number of nodes in the network, i. e. $\Delta=N$. However, if node degree is selected as the weighting, i.e. BFCM-Degree algorithm, then

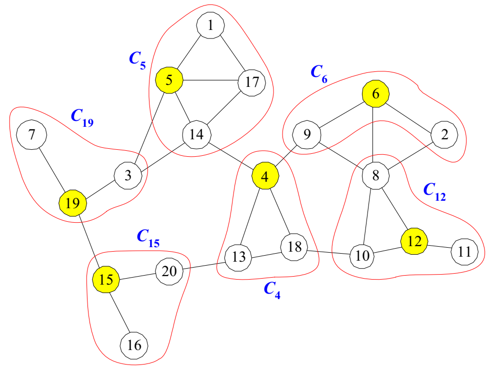

Figure 2 A connected ad hoc network with 20 nodes. Numbers in circles are node IDs. For BFCM-Degree algorithm, 6 clusters are generated and circles with shaded area are the corresponding clusterheads.

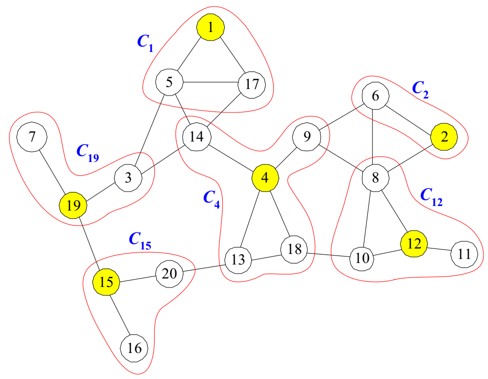

Figure 3 A connected ad hoc network with 20 nodes. For BFCM-ID algorithm, 6 clusters are generated and circles with shaded area are the corresponding clusterheads.

$f(v)=\deg(v)$ and Δ may be set to the maximum possible node degree for an N nodes connected network, i. e. $\Delta=N-1$ (a star topology).

C. *Distributed Boundary-First Cluster-Minimized (BFCM) Clustering Algorithm*

Following, based on the assumption that each node knows the IDs and degrees of all one-hop neighbors, we provide the distributed BFCM clustering algorithm. The operation of the algorithm is stated as follows and is concurrently executed in each unmarked node v:

1. Regards itself as a clusterhead, updates status to marked and broadcasts an invite packet, **Invite(*v*)**, to all neighbors if
   1.1. Node v is a typical node, or
   1.2. Node v is the only node with maximum weight among unmarked neighbors, or
   1.3. Among unmarked neighbors with the same maximum weight, node v with the smallest ID.
2. On receiving an invite packet sent from neighboring node u,

2.1. If node v is a typical node, discards this packet.

2.2. Otherwise, regards itself as an ordinary node, updates status to marked and sends a **Join(v,u)** packet to join the cluster constructed by node u. (If more than one invite packet received, selects the sender with the largest weight. If there are more than two such senders, selects the one with the smallest ID.)

Example: Consider a 20 nodes connected ad hoc network with three boundary nodes, node 11, 16 and 7. The three corresponding typical nodes are node 12, 15 and 19, respectively, as illustrated in Figure 2. If the BFCM-Degree clustering algorithm is employed, Δ=19 and there are 6 clusters generated, C_4, C_5, C_6, C_{12}, C_{15}, C_{19}, as shown in Figure 2. The subscript indicates the clusterhead ID. If the BFCM-ID clustering algorithm is used, Δ=20 and there are also 6 clusters, C_1, C_2, C_4, C_{12}, C_{15}, C_{19}, generated as shown in Figure 3. Obviously, there is no orphan cluster generated in Figure 2 and Figure 3. If the Degree-based [5] and ID-based [6]-[7] algorithms are used to the same network, there are 6 clusters and 8 clusters generated respectively. However, among the generated clusters, there are 2 orphan clusters created by boundary nodes.

To prove the correctness of the proposed algorithm, we need to show that the generated cluster satisfies the dominance and independence constraints and the algorithm terminates eventually.

Property 1: Every ordinary node has at least a clusterhead as neighbor. (*dominance constraint)*

Proof: It is easy to show that for any ordinary node, it must exist a clusterhead as neighbor (otherwise, it is a clusterhead). Consider the scenario as shown in Figure 5. A marked ordinary node c with clusterhead node d has at least one unmarked node b whose weight is local maximum among its unmarked neighbors. Then, node b will be elected as a clusterhead to organize a new cluster. As a consequence, the ordinary node c has two clusterheads as neighbors.

Property 2: The generated clusters would not mutually overlap. (independence constraint)

Proof: According to the procedures of the BFCM algorithm, whenever a node is clustered, it updates node status from unmarked to marked; moreover, only unmarked nodes execute the algorithm. Hence, a marked node is never being processed again, i.e., a node would not belong to more than one cluster. This proves no clusters are mutually overlapped.

Property 3: The aggregate number of nodes in each cluster is the total number of nodes in the network.

Proof: Let nodes in cluster C_{m_i} are represented as a set V_{m_i} where m_i is the clusterhead ID of the corresponding cluster. Assume the algorithm generates i clusters, C_{m_1}, C_{m_2} ,…, C_{m_i} . From Property 2, we know that each set V_{m_k} with respect to cluster C_{m_k} is disjointed. According to the set operation, we

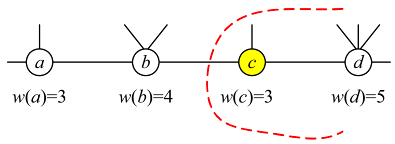

Figure 5 Ordinary node c contains two clusterheads as neighbors.

have $\bigcup_k V_{m_k} = V$ where k=1, 2,…, i. Thus, we have $\sum_k \left|V_{m_k}\right| = \left|V\right| = N$.

Property 4: The algorithm terminates in finite steps.

Proof: Since only unmarked nodes are considered in this algorithm and whenever a node is clustered, its status will be

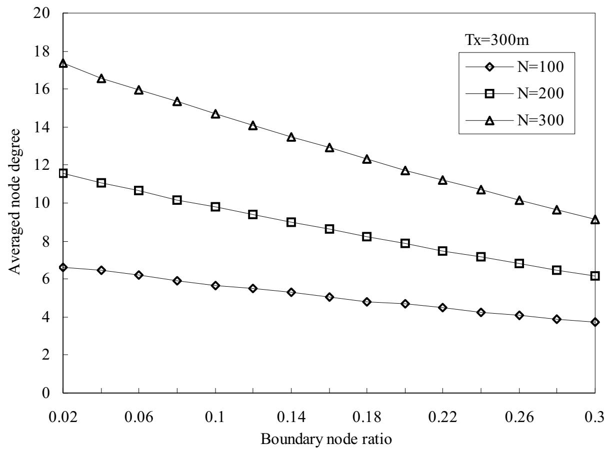

Figure 4 The averaged node degree of the generated network.

updated to marked. It is obvious that the algorithm terminates when the unmarked node set is empty.

Property 5: The message complexity of the proposed algorithm is $O(N)$.

Proof: Since events that trigger transmitting of **Invite(v)** and **Join(v,u)** packets are mutually exclusive, each node will only broadcast one packet when execution the algorithm terminates. Thus, the message complexity is $O(N)$.

Property 6: The time complexity of the proposed algorithm is $O(N)$.

Proof: Since there are totally N nodes in the network and from *Property 5*, each node transmits only one packet and each packet is processed by constant number of steps. Thus, we can easily to show that the time complexity is $O(N)$.

IV. SIMULATION AND RESULTS

We verify the effectiveness of the proposed algorithm by conducting extensive simulations and compare the obtained results to the Degree-based and ID-based [5]-[7] algorithms. Each simulation result in the following figures is the averaged value of 10,000 simulations. In each simulation, we generate a connected ad hoc network by randomly placing N nodes in a 2000m×2000m square. The transmission range of each node is

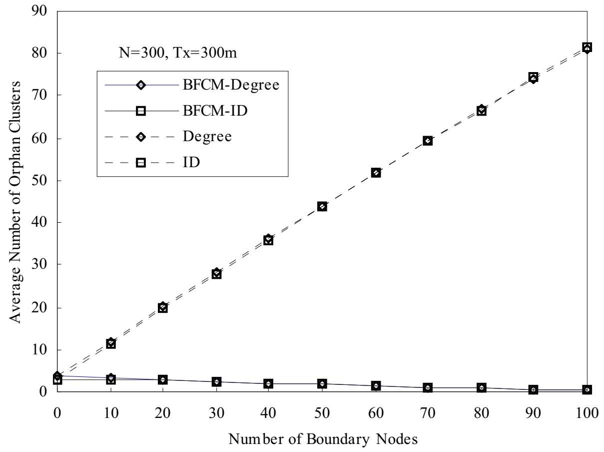

Figure 6 Comparison of average number of orphans.

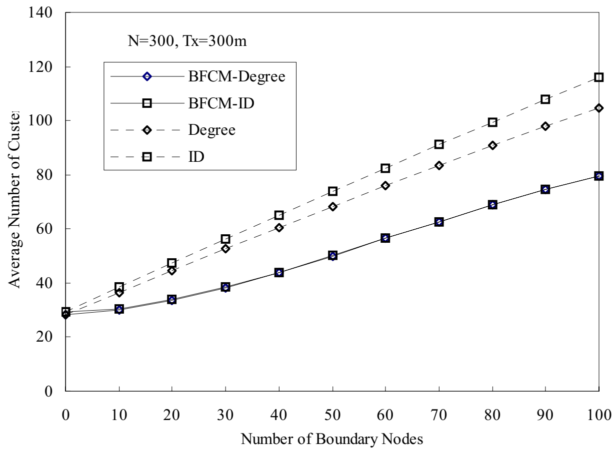

Figure 7 Comparison average of number of clusters.

300m and is fixed. The first result shown in Figure 4 reveals that boundary node reduces the average network node degree. This means that the network is getting "*sparse*". Besides, it also shows that the ratios of the averaged node degree are approximately proportional to the ratios of number of node. For example, for N=100 and 300, their corresponding slopes are S_{100} = -10.12 and S_{300} = -32.84 $\cong$ 3 S_{100}.

Figure 6 compares the number of orphan cluster generated by the proposed algorithm to the other two algorithms. It shows that the proposed algorithms reduce the number of generated orphan clusters dramatically. This result also shows that assigning typical nodes with the highest weighting, Δ, in (1) is a good way to reduce the number of orphan cluster generated by boundary nodes. Figure 7 shows that the averaged total number of organized clusters by the proposed algorithm is far less than the others. In addition to evaluate performance of the BFCM algorithm in the *averaged sense* as shown in Figure 6 and Figure 7, we also strictly to probe each of the simulation results to verify if there is any violation to the comments for Figure 6 and Figure 7 occurs. Perfectly, the probing results indicate that NO violation occurs during the simulations!!

V. Conclusions

To reduce communication overheads in transmitting routing information and network maintenance message, incorporating with a generalized weighting function, we present a distributed BFCM algorithm to minimize the total number of generated clusters in an ad hoc network. We first transform this cluster minimization problem into selection of the nodes with the most number of neighbors as clusterheads to maximize the members in a cluster; and, then, based on the characteristic that boundary nodes is the main reason to generate orphan cluster, we further minimized the generation of orphan clusters. Through complexity analyses and simulations, we conclude that with remaining the same message and time complexities, $O(N)$, to the compared ID-based and Degree-based algorithms, the proposed BFCM algorithm combining with a generalized weighting function is effective in minimizing the orphan clusters generated by the boundary nodes and reducing the total number of generated clusters not only in the averaged sense but also in the strict sense.

Reference

[1] B. Das and V. Bhargavan, "Routing in ad-hoc networks using minimum connected dominating sets," *IEEE ICC 97*, 1997.

[2] J. Wu and H.L. Li, "On calculating connected dominating set for efficient routing in ad hoc wireless networks," *ACM Proceedings of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pp. 7-14, 1999.

[3] M. R. Garey and D. S. Johnson, "Computers and Intractability: A guide to the theory of NP-Completeness," *Freeman*, *San Francisco*, 1978.

[4] C.S. Li, "Clustering in packet radio networks," *IEEE ICC 85*, pp. 283-287, 1985.

[5] M. Gerla and J. T. C. Tsai, "Multicluster, mobile, multimedia radio network," *ACM/Baltzer J. Wireless Networks*, vol. 1, pp. 255-265, 1995.

[6] D. J. Baker and A. Ephremides, "The architectural organization of a mobile radio network via a distributed algorithm," *IEEE Trans. on Commu.*, vol. COM-29, no. 11, pp. 1694-1701, Nov. 1981.

[7] C.H. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE Journal on Selected Area of Communications*, vol. 15, no. 7, pp. 1265-1275, 1997.

[8] M. J. Shah and P. G. Flikkema, "Power-based leader selection in ad-hoc wireless networks," *IEEE International Performance, Computing and Communications Conference*, 1999.

[9] S. Singh and J. Kurose, "Electing leaders based upon performance: the delay model," *IEEE International Conference on Distributed Computing,* pp. 464-471, 1991.

[10] S. Basagni, "Distributed clustering for ad hoc networks," *Proceedings of the 1999 International Symposium on Parallel Architectures, Algorithms, and Networks (I-SPAN'99)*, pp. 310-315, 1999.

[11] P. Basu, N, Khan and Thomas D. C. Little, "A mobility based Metric for clustering in mobile ad hoc networks," *IEEE International Conference on Distributed Computing Systems Workshop*, pp. 413-418, 2001.

[12] Mario Joa-Ng and I-Tai Lou, "A peer-to-peer zone-based two-level link state routing for mobile ad hoc networks," *IEEE Journal on Selected Area of Communications*, vol. 17, no. 8, pp. 1415-1425, 1999